

# Symbiotic gut microbes modulate human metabolic phenotypes

Min Li\*, Baohong Wang<sup>†</sup>, Menghui Zhang\*, Mattias Rantalainen<sup>‡</sup>, Shengyue Wang<sup>§</sup>, Haokui Zhou\*, Yan Zhang\*, Jian Shen\*, Xiaoyan Pang\*, Meiling Zhang\*, Hua Wei\*, Yu Chen<sup>†</sup>, Haifeng Lu<sup>†</sup>, Jian Zuo<sup>†</sup>, Mingming Su\*, Yunping Qiu\*, Wei Jia\*, Chaoni Xiao<sup>||</sup>, Leon M. Smith<sup>‡</sup>, Shengli Yang\*, Elaine Holmes<sup>‡</sup>, Huiru Tang<sup>||\*\*</sup>, Guoping Zhao<sup>§\*\*</sup>, Jeremy K. Nicholson<sup>‡\*\*</sup>, Lanjuan Li<sup>†\*\*</sup>, and Liping Zhao<sup>\*\*\*</sup>

\*Ministry of Education Key Laboratory of Systems Biomedicine, Shanghai Center for Systems Biomedicine at Shanghai Jiao Tong University, Shanghai 200240, China; <sup>†</sup>State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University, Hangzhou 310003, China; <sup>‡</sup>State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan 430071, China; <sup>§</sup>Graduate University of the Chinese Academy of Sciences, Beijing 100049, China; <sup>||</sup>Shanghai-MOST Key Laboratory for Disease and Health Genomics, Chinese National Human Genome Center, Shanghai 201203, China; and <sup>\*\*</sup>Department of Biomolecular Medicine, Faculty of Medicine, Imperial College London, South Kensington SW7 2AZ, United Kingdom

Communicated by Zhu Chen, Shanghai Second Medical University, Shanghai, People's Republic of China, December 22, 2007 (received for review October 27, 2007)

Humans have evolved intimate symbiotic relationships with a consortium of gut microbes (microbiome) and individual variations in the microbiome influence host health, may be implicated in disease etiology, and affect drug metabolism, toxicity, and efficacy. However, the molecular basis of these microbe–host interactions and the roles of individual bacterial species are obscure. We now demonstrate a “transgenomic” approach to link gut microbiome and metabolic phenotype (metabotype) variation. We have used a combination of spectroscopic, microbiomic, and multivariate statistical tools to analyze fecal and urinary samples from seven Chinese individuals (sampled twice) and to model the microbial–host metabolic connectivities. At the species level, we found structural differences in the Chinese family gut microbiomes and those reported for American volunteers, which is consistent with population microbial cometary differences reported in epidemiological studies. We also introduce the concept of functional metagenomics, defined as “the characterization of key functional members of the microbiome that most influence host metabolism and hence health.” For example, *Faecalibacterium prausnitzii* population variation is associated with modulation of eight urinary metabolites of diverse structure, indicating that this species is a highly functionally active member of the microbiome, influencing numerous host pathways. Other species were identified showing different and varied metabolic interactions. Our approach for understanding the dynamic basis of host–microbiome symbiosis provides a foundation for the development of functional metagenomics as a probe of systemic effects of drugs and diet that are of relevance to personal and public health care solutions.

covariation analysis | gut microbiota | metabonomics | metabotype | metagenomics

Human beings can be considered as “superorganisms” as a result of their close symbiotic associations with the gut microbiota (1). Superorganism metabolism involves integration of truly indigenous metabolic processes (coded in the host genome) with those of the microbiome. This results in extensive transgenomic cometary metabolism of many substrates including those involved in host metabolic regulation (2). The superorganism concept represents an important paradigm shift in understanding human biology and is likely to have a significant impact on the future of disease prevention and therapy (3). Recent works have shown that the exact human microbiome composition varies between healthy people (2, 4, 5) and also between lean and obese individuals (6), and further, that the microbiome composition is responsive to dietary modulation for weight reduction (6). “Top-down” systems biology (3) analysis of metabolic profiles of human baby microbiota and normal microbiota associated mice revealed that absorption, storage, and metabolism of dietary

lipids were specifically modulated by the microbiome (7). Moreover, the induction of type 2 diabetes and obesity with a high-fat diet in rats has been shown to correlate with the pre-dose metabolic patterns associated with differences in gut bacterial activities, indicating the importance of the microbiome in host predisposition to diseases (8). Recent work showed that gut microbiome was probably responsible in part for the systemic response to *Schistosoma mansoni* infection in mice (9). Disruptions of choline metabolism caused by changes in symbiotic gut microbiota may play an active role in the development of insulin resistance and nonalcoholic fatty liver disease in high-fat diet experiments with a mouse strain genetically predisposed to these disease traits (10). Responses of individual animals and humans to drug treatments can also be strongly influenced by gut microbiome composition, because the microbiome provides not only complementary metabolic pathways for drugs, but is also a source of pharmacologically active secondary metabolites that can activate mammalian liver enzyme systems (2, 3). The importance of gut microbiota to host metabolism may best be illustrated by the fact that genetically homogeneous animals can have diverse metabolic phenotypes when they have structurally different gut microbiota (2). It has also been shown that identical twins still had significant differences in their gut microbiota although they shared much higher similarity for gut microbiota structures than genetically unrelated married couples (11). The unique combination of gut bacteria in each animal may thus have an important role in their host's metabolism because they are adding new dimensions to functional diversity at the whole-organism level on host genetics, which includes participating the development of pathophysiology (10) and providing complementary metabolic pathways for drugs and diets (2). In light of the recent findings on the relationships between gut bacterial composition and the obese host phenotype (12), understanding gut bacterial dynamics in relation to host physiology and pathology

Author contributions: M.L., B.W., Menghui Zhang, M.R., and S.W. contributed equally to this work; W.J., S.Y., H.T., G.-P.Z., J.K.N., L.L., and L.-P.Z. designed research; M.L., B.W., Menghui Zhang, M.R., S.W., J.S., X.P., Meiling Zhang, H.W., Y.C., H.L., J.Z., Y.Q., and C.X. performed research; H.T., J.K.N., and L.-P.Z. contributed new reagents/analytic tools; M.L., B.W., Menghui Zhang, M.R., S.W., H.Z., Y.Z., M.S., L.M.S., S.Y., E.H., H.T., G.-P.Z., J.K.N., L.L., and L.-P.Z. analyzed data; and M.L., B.W., Menghui Zhang, M.R., S.W., E.H., H.T., G.-P.Z., J.K.N., L.L., and L.-P.Z. wrote the paper.

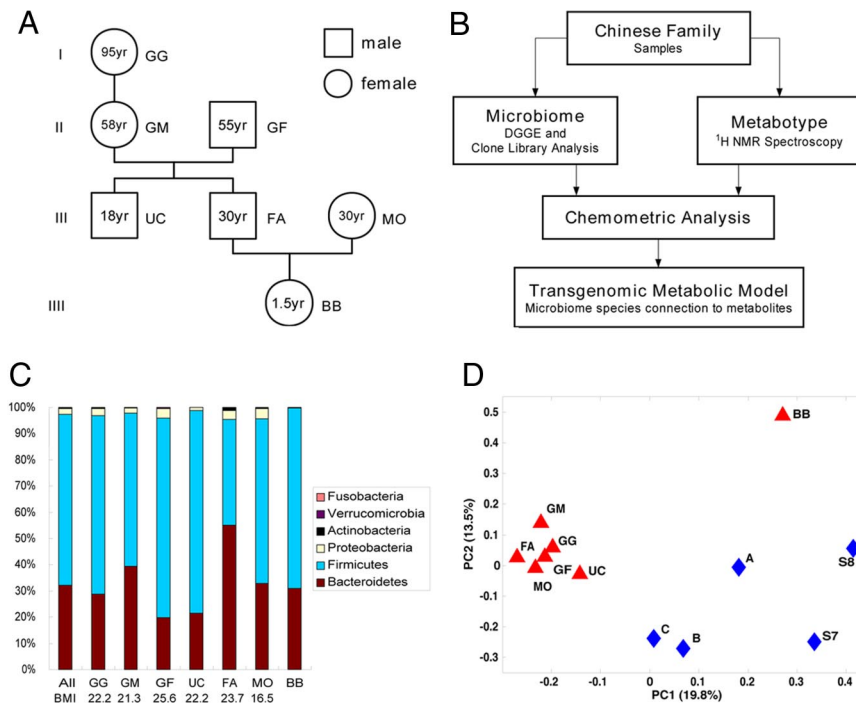
The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EF398274–EF405528 and EF395822–EF395843).

\*\*To whom correspondence may be addressed. E-mail: huiru.tang@wipm.ac.cn, gpzhao@sibs.ac.cn, j.nicholson@imperial.ac.uk, ljli@zju.edu.cn, or lpzhao@sjtu.edu.cn.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0712038105/DC1](http://www.pnas.org/cgi/content/full/0712038105/DC1).

© 2008 by The National Academy of Sciences of the USA



**Fig. 1.** Experimental procedure and structural comparison of gut microbiome between Chinese and American individuals. (A) Family tree diagram of the Chinese family. (B) Scheme of experimental procedure. (C) The division-level composition of gut microbiome of the Chinese family. (D) Species-level composition of gut microbiome of the Chinese family in comparison with reported American microbiome data (4, 5). The principal coordinate scores plot was generated by using UniFrac metrics. The percentages of variation described by the principal coordinates are shown in the parentheses.

has become an important part of future personalized health care solutions (3, 13).

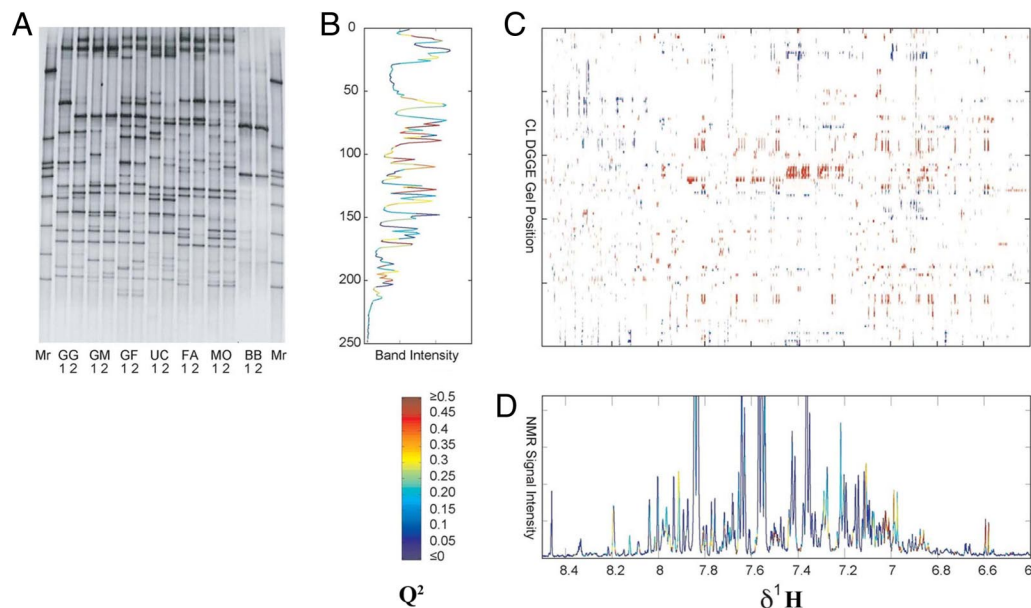
Although the importance of the gut microbiome in host health is now widely recognized, it is not yet known which of the many hundreds of species are of “key” importance in host health, and little is understood of the molecular host–microbiome interactions that influence host metabolic pathways. The chemical signatures of biofluids can provide important metabolic phenotype (metabotype) (14) information that characterizes organismal level metabolic status and is reflective of symbiotic metabolic complexity and cometabolism (3). Thus, we postulate that covariation between the metabolic phenotype and the gut microbiome structure may be exploited to identify specific associations between key members of the gut microbial community and patterns in host physiology or pathology. Most gut microbial metabolism analytical surveys have typically focused on single or targeted components of the system, for example, short-chain fatty acids (12), and have not addressed the wider complexity of host–microbiome metabolic interactions in superorganisms. So here we have developed a multivariate strategy to model covariation between gut microbiomic structural patterns as reflected by community DNA fingerprints and host metabotype as defined by NMR spectroscopic urinary profiling (15). Patterns of covariation are extracted by using multivariate statistical approaches to establish the association between host metabolism and variation of gut microbiota. This approach allows the visualization and mapping of the transgenomic interactions of microbiome and host by the noninvasively obtained excreted metabolite profile and fecal microbial composition, which holds promise for functional characterization of gut microbiota and for understanding human health and disease at both the individual and the population level.

## Results

**Phylogenetic Landscapes of Gut Microbiota in the Chinese Family.** A clone library approach was used first to investigate the diversity

of gut microbiomes in the cohort to obtain a detailed structural overview of the microbiome of each member of the study family. The clone library of each individual was shown to have >93% Good’s coverage, indicating that the 16S rRNA gene sequences from these samples represented the majority of human intestinal bacterial community in this study [supporting information (SI) Table 1]. Phylogenetic analysis of the total 7,255 16S rRNA gene sequences from the seven individuals yielded 476 operational taxonomic units (OTUs) with a 99% similarity cutoff (SI Table 1 and SI Figs. 4 and 5). The overall (SI Fig. 6) and individual (Fig. 1C) structures of the microbiome were dominated by *Firmicutes* and *Bacteroidetes* as found in earlier studies (4). *J*-LIBSHUFF analysis (16) also showed that each individual appears to have a unique gut microbial community even within this one family (details in SI Text: Results) in agreement with previous reports about the individuality of human gut microbiota. The relative abundance of the major bacterial divisions (Fig. 1C) shows a high degree of interpersonal variation in the *Bacteroidetes*-to-*Firmicutes* ratios ranging from 0.26 to 1.36. It was recently shown that a low *Bacteroidetes*-to-*Firmicutes* ratio was correlated with obesity and that this ratio can be increased by dietary calorific restriction (6). All of the Chinese studied here had *Bacteroidetes*-to-*Firmicutes* ratios that were similar to lean American individuals reported in previous studies (6). However, in our study, the only marginally overweight family member [GF, body mass index (BMI) 25.6] also had the lowest *Bacteroidetes*-to-*Firmicutes* ratio (0.26) in his gut microbiota. Interestingly, the other family member (UC) who had a low *Bacteroidetes*-to-*Firmicutes* ratio (0.28) was not overweight, but had lived in the United Kingdom for 2 years and adopted a more Western lifestyle and diet pattern (details in SI Text: Materials and Methods).

When compared with data obtained from two other recent studies conducted on American volunteers by using the same genetic microbiological tools (4, 5), the Chinese family studied here shared a similar division-level phylogenetic landscape with



**Fig. 2.** Multivariate analysis for identifying associations between the gut microbiome structure and the urine metabolite profile. (A) DGGE gel for *C. leptum* subgroup. Mr, marker lane. (B) OPLS prediction of clostridia bands from the NMR urinary profile data. (C) Two-dimensional correlation map of NMR-derived metabolic profile variation in relation to DGGE fingerprints (only the aromatic region of urinary NMR spectra is shown); only points with absolute correlation level  $>0.7$  are shown, red denotes positive correlation, blue denotes negative correlation. (D) OPLS prediction of aromatic region of the NMR spectrum from DGGE data. The color indicates the  $Q^2$  value.

the American individuals (SI Fig. 6). However, the “Chinese” and “American” microbiomes were clearly different at the species level of composition as shown by the principal coordinate analysis scores plot based on UniFrac metrics (17) (Fig. 1D), which also illustrated the relative similarity of the Chinese adults, and the large differences in the baby microbiota profile. Because the Chinese baby had an immature and unique microbiome structure compared with the adults (Fig. 1D), the baby samples were excluded from the next level of metabolic linkage analysis.

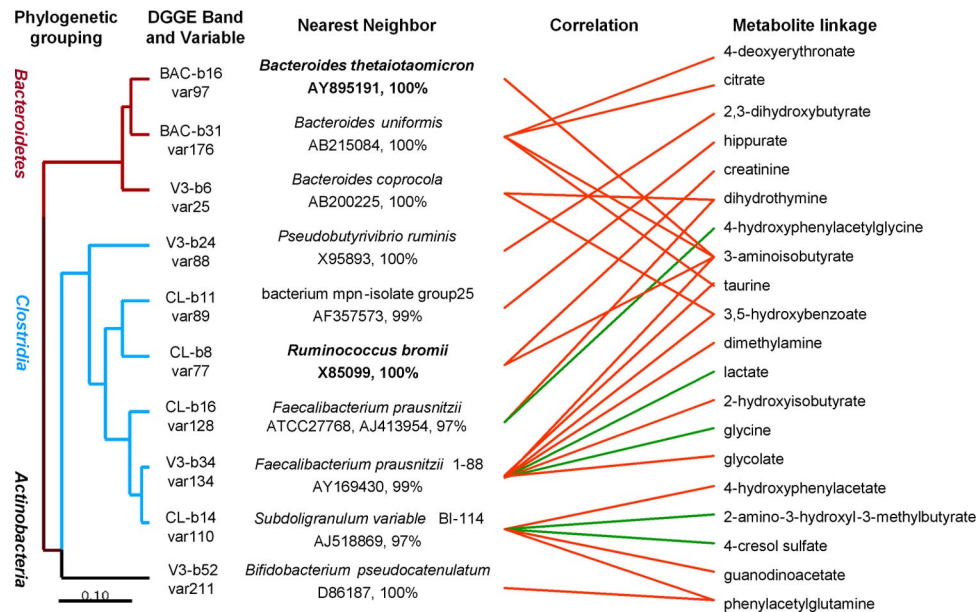
**Denaturing Gradient Gel Electrophoresis (DGGE) Profiling of Inter- and Intraindividual Variation in Gut Microbiota of the Chinese Family.** After the overall structural survey of the Chinese family microbiome, the short-term time stability of the overall structure of the microbiome over a month within the cohort was illustrated by DGGE patterns of the 16S rRNA gene V3 region for predominant bacteria and of regions for two specific groups, *Bacteroides* spp. and *Clostridium leptum* subgroup (see SI Figs. 9B and 10A and Fig. 2A). The analysis of the microbiota DGGE fingerprints by using principle components analysis (PCA) and fivefold cross-validated orthogonal projection to latent structures discriminant analysis (OPLS-DA) (18) showed a sex-related difference for all of the three types of DGGE patterns (SI Figs. 7 and 8). The key band variables, which were responsible for the separation, were identified from the DGGE gels by sequencing (SI Table 2). Among these sex-related bacteria, three species were from *Clostridia*, one from *Bacteroidetes*, and two from *Proteobacteria*. All of these species had higher abundance in males than in females.

**Metabolic Profiling of Metatypes Variation by <sup>1</sup>H NMR Spectroscopy.** For metabolic profiling of the urine samples, 600-MHz <sup>1</sup>H NMR spectroscopy was performed. PCA and fivefold cross-validated OPLS-DA applied to the full NMR spectra data also showed the suspected gender-related difference in urine metabolites of family members (SI Figs. 7D and 8G) that was consistent with the PCA analysis of gut bacteria DGGE profiling. The important gender-related variables for the discrimination were

statistically and structurally identified as 3-aminoisobutyrate (BAIB, doublet  $\delta 1.2$ ), which was higher in males, and creatine (singlet  $\delta 3.05$ ), which was higher in females.

**Covariation Analysis of DGGE Microbiota Profiling and NMR-Based Metabolic Profiling.** Covariation between NMR urine spectra and DGGE gel data were modeled by using OPLS regression (18). The predictive power of the model was assessed by fivefold cross-validation, and the  $Q^2$  value (goodness of prediction) was color-mapped onto the DGGE banding pattern, showing how well each band was predicted by the spectroscopic data. The well predicted bands were then sequenced and identified as OTUs based on their nearest neighbors in the RDP-II database. Conversely, the regions of the NMR spectra that were best predicted by the DGGE data were color-coded (according to  $Q^2$ ) onto the NMR profile. The DGGE gel for the *C. leptum* subgroup was used to exemplify this analysis strategy (Fig. 2). The methodology was first applied to the DGGE gel fingerprinting of the predominant bacteria. The bands, which were best predicted by the variation in the NMR data, are identified as OTUs in *Bacteroidetes*, *Firmicutes*, and *Actinobacteria* (SI Fig. 9). The closest relative isolates of these OTUs are mostly well known gut bacteria, such as *Bacteroides coprocola*, *Faecalibacterium prausnitzii*, and *Bifidobacterium pseudocatenulatum*. Having established the predominant bacteria, we repeated the analysis by using DGGE fingerprints for specific groups of bacteria to identify key OTUs at a deeper level. More OTUs were identified in each group in addition to those identified from the V3 gel (Fig. 2 and SI Fig. 10). For instance, in the *C. leptum* subgroup, some more OTUs such as *Ruminococcus bromii* and *Subdoligranulum variable* were predicted in group-specific DGGE gel, in addition to *F. prausnitzii*, whereas in the *Bacteroides* spp.-specific DGGE gel, *Bacteroidetes thetaiotaomicron* and *Bacteroidetes uniformis* were identified in addition to *B. coprocola*.

The relationships between the metabolic data and the microbiome data were visualized in the form of transgenomic cross-correlation maps, which displayed both positive and negative (e.g.,  $r > 0.7$  and  $< -0.7$ ) correlations between the NMR and DGGE



**Fig. 3.** Dendrogram of OTUs from DGGE bands, which are well predicted by metabolic variation, labeled as the nearest known neighbor with similarity value. Associations with specific urine metabolites are shown for each OTU with the direction of correlation indicated by red (positive) or green (negative) lines. Gender-related bands predicted by OPLS-DA are denoted by bold text.

data matrices (Fig. 2C). The metabolic associations of each of the well predicted members of the gut microbiome were shown in Fig. 3. Many of the identified compounds are well known gut microbial cometabolites such as phenylacetylglutamine, 4-cresol sulfate, and 4-hydroxyphenylacetate (SI Table 3).

Some microbial members only reveal a single metabolite association, for example, the connection between *B. thetaiotaomicron* and 3-aminoisobutyrate which also both correlated with sex, other members have multiple connectivities, for example *F. prausnitzii* was statistically linked with the presence of 8 urinary metabolites including dimethylamine, taurine, lactate, glycine, 2-hydroxyisobutyrate, glycolate, 3,5-hydroxybenzoate, and 3-aminoisobutyrate.

## Discussion

Gut microbial composition among healthy people is influenced by host genotype, diet, age, and sex, and it appears that organic disease and drugs (especially antibiotics) can modulate microbiome composition and activities (2, 3). The observed host metabolic phenotype is thus strongly influenced by the gut microbiome (2). Host genetics affects the broad gut microbial structure as evidenced by the widely differing species compositions of various animal species. On short time scales, the host-specific effects are relatively constant and changes in the gut microbiota and host metabolism will be closely influenced by dietary variation. To understand these interactions, it is essential to understand which host metabolic pathways are most closely associated with structural variation of gut microbiota and vice versa.

In the Chinese family cohort studied here, background data on the genetic relatedness among members, the sex, age, and dietary pattern of each individual, are relatively more defined than a random population cohort. Sampling at two time points increases statistical power and with appropriate multivariate methods it is possible to dissect the relationships between gut microbial variation and the interacting host metabolic pathways and assessing cross-sectional and longitudinal variations.

The overall species composition of gut microbiota is significantly different between this Chinese family and previously reported Americans (4, 5). This is coherent with the recent

epidemiological observation that metabolic phenotypes of the Chinese population differed from the metabolic phenotypes of the American population (15), and the gut microbial-mammalian cometabolites were distinguishing biomarkers, including hippurate and phenylacetylglutamine.

A previous cross-sectional study ( $n = 230$  from four European countries) also showed that gender-specific differences were observed in the *Bacteroides-Prevotella* group with males more prominent than females (19). However, the gender-correlated key species were not identified probably because only species-specific probes for *Bacteroides vulgatus* and *Bacteroides putredinis* were used and neither showed sex-related difference. By using our *Bacteroides* spp. group-specific DGGE profiling method, we identified *B. thetaiotaomicron* as being higher in males than females. Variation of the *B. thetaiotaomicron* population level was positively correlated with urinary BAIB, which was also a sex discriminator for Chinese (15).

By correlating the DGGE pattern with the urinary NMR profiles, the covariance of structural variation in gut microbiome and host metabolism were visualized in the current work (Fig. 2). The OTUs of DGGE variables, which were best predicted by the variation in the metabotype data, were related to *Bacteroides*, *Clostridium*, and *Bifidobacteria*. Isolates of the species nearest to these identified OTUs have been extensively studied with respect to interactions with host energy bioavailability and immunity (20, 21). *F. prausnitzii* is the most significant *n*-butyrate-producing gut bacterium with well known effects on host energy metabolism and mucosal integrity (22). *B. pseudocatenulatum* is well known for its conjugated linoleic acid (CLA) biosynthetic ability which has anti-carcinogenic effects (23). The identification of metabolically linked OTUs related to these known important members of the gut microbiota adds weight to the utility of this new methodology.

Conversely, the regions of the NMR spectra, which were well predicted by the individual DGGE OTUs, had been structurally identified as the well known metabolites produced by gut bacteria or cometabolites with the host, such as phenylacetylglutamine, 4-cresol sulfate, and 4-hydroxyphenylacetate. 4-Cresol sulfate has previously been identified as mammalian metabolite of 4-cresol that can be synthesized by only a few

species of bacteria including *Clostridium difficile* (24). Here, we also observed a statistical association between 4-cresol sulfate and *Subdoligranulum variable* BI-114, which is genetically closely related to the clostridia. The negative association between *S. variable* and 4-cresol sulfate may indicate a competitive relationship between this bacterium and its close relatives that can produce 4-cresol. Variation of these metabolites in urine has been used to indicate the importance of gut microbial dynamics over time or changes from individual to individual in disease progression or drug metabolism (2). Identification of these gut bacterial metabolites confirmed the power of this method to link specific metabolite with a specific species or group of gut bacteria.

The number and diversity of metabolic pathways associated with OTUs may reflect the magnitude of importance of these members of the gut microbiome in host physiology. Although some members only reveal a single metabolite association, other members have multiple connectivities. These data suggest that there are profound host–microbiota symbiotic connections that influence global metabolism regardless of the genetic background across a range of pathways or environmental exposure of the host. The connectivity between dimethylamine and *F. prausnitzii* is particularly intriguing because dimethylamine is a product of microbial catabolism of dietary choline (by trimethylamine) that has previously been linked to high-fat diet-induced insulin resistance, fatty liver, and type 2 diabetes in experimental mice (10), and in other studies a low *Bacteroidetes*-to-*Firmicutes* ratio was identified in both genetic models of obesity (12) and in obese people (6). Furthermore, we have shown that gut microbial variation in mice strongly influences bile acid metabolism, including higher enterohepatic recycling of taurine conjugates that are important in the emulsification of fats (7) (and therefore lipid bioavailability to the host), and the connectivity found here to free urinary taurine levels for *F. prausnitzii* and *B. uniformis* may also reflect this interaction. This provides further indirect evidence of the connectivity between disturbances in the gut microbial populations and the metabolic consequences of the altered microbial–mammalian metabolic balance influencing host disease.

We have presented a proof-of-principle for the development of a transgenomic methodology for exploring microbial host metabolic relationships. The clone library was not used in this work for covariation analysis because of its qualitative nature and the uncertainty of its reproducibility (25–27). However, with the rapid development of high-throughput sequencing techniques (12, 28), sequencing-based analysis will eventually become quantitative and cost-effective enough for this type of analysis. We can even anticipate a covariation analysis between host metabolites and composition of a gut bacterial functional gene pool obtained by low-cost, high-throughput, nonbiased, and quantitative sequencing and annotation of whole-gut microbiota DNA from multiple samples with statistical power. Thus, a molecular link between host metabolism and gut microbiota genes, rather than just species identity, can be established in the end for more profound understanding of metabolic basis of host health and their responses to various drug/diet modulations.

In conclusion, this approach allows identification of potentially important associations between changes of bacterial community structure and dynamics of host metabolic patterns that can be used to develop new and existing hypotheses on the relationships between dysbiosis and disease. This approach should lay a foundation for the development of *functional metagenomics*, which is currently constrained by the difficulties in studying the functional ecology of the symbionts *in situ*. Specifically, we can now identify microbial candidates for sequencing that have strong host metabolic connectivities, because these are likely to be more relevant to host biology and health. Thus, this work is a step in the quest for the “Rosetta stone” that

will allow translation of microbial community activity into host response, which, in turn, will enable an understanding of extended genome-related disease processes and may provide a means of engineering metabolic activities of gut microorganisms to better suit human health.

## Materials and Methods

**Subject Selection and Sampling.** A volunteer four-generation Chinese family (Fig. 1A) including three male (ages, 18–55 years) and four female (ages, 1.5–95 years) members living in three separate households, two in Hangzhou (GG, GM, and GF in one household; FA, MO, and BB in another), China, and one (UC) in London was chosen for study. The use of these subjects was approved by the First Affiliated Hospital of Zhejiang University Institutional Review Board. All participants in this study provided their informed consent. Fecal and early-morning urinary samples from each individual were collected on two occasions, 30 days apart. All samples were immediately frozen on collection and stored at  $-70^{\circ}\text{C}$  before analysis.

**Phylogenetic Analysis of Large-Scale 16S rRNA Gene Sequencing Data.** One clone library was constructed for each family member with one time point sample by PCR amplification of near full-length 16S rRNA genes with primer P0 (*Escherichia coli* position 7–27) and 1492r (*E. coli* position 1492–1511) (29, 30). An average 1,000 positive clones of each subject were picked randomly for sequencing in both directions with ABI 3730xl sequencers (Applied Biosystems). The manually checked high-quality reads with average length of 750 bp were assembled into consensus sequences by the program CodonCode Aligner (CodonCode Corporation). The assembled sequences were first trimmed to exclude vector sequences, and then checked for chimeras by using CCODE (31) and Chimera Check v2.7 on Ribosomal Database Project II (RDP-II) website (<http://rdp.cme.msu.edu/cgiis/chimera.cgi?su=SSU>). Finally, all full-length 16S rRNA sequences were aligned to the small subunit rRNA sequences in the RDP-II (32) by using ARB software package (33). A phylogenetic tree was generated by a neighbor-joining algorithm from an Olsen-corrected distance matrix. Sequences were grouped into OTUs at the 99% similarity threshold from Olsen-corrected distance metrics by DOTUR with the furthest-neighbor algorithm and 0.001 precision (34).

To find the population difference between gut microbial communities, we compared the clone library data of the Chinese family with the data of five American individuals from two recent studies (4, 5) by using  $\beta$ -LIBSHUFF analysis (16) and the UniFrac method (17) (details in *SI Text: Materials and Methods*).

**DGGE Analysis of Fecal Samples.** The V3 regions of 16S rRNA genes and specific DNA fragments of two predominant bacterial groups in the gut (*Bacteroides* spp., *C. leptum* subgroup) from fecal samples of two time points were amplified by universal bacterial primers and group-specific primers (35–37). PCR products were separated in 8% (wt/vol) denatured polyacrylamide gels by electrophoresis. The DGGE image was read by ImageJ software (<http://rsb.info.nih.gov/ij/>) and the intensity and position of bands in each lane were read into a spectrum of 249 variables. The spectra were manually corrected for gel shift (details in *SI Text: Materials and Methods*).

Selected DGGE bands were excised from original gels and their DNA fragments were reamplified with corresponding primers. The PCR products were sequenced as described earlier. The sequences were submitted to RDP-II release 9 database to determine their closest isolate relatives. The phylogenetic tree was constructed based on sequences of their closest neighbors of DGGE bands.

**NMR Spectroscopic Analysis of Urinary Samples.** An aliquot (320  $\mu\text{l}$ ) of each urine sample was mixed with 160  $\mu\text{l}$  of sodium phosphate buffer (0.2 M, pH 7.4) and 120  $\mu\text{l}$  of deuterium oxide containing 0.05% (wt/vol) 3-trimethylsilyl-1-[2,2,3,3- $^2\text{H}_4$ ] propionate (TSP) as a reference. All NMR analysis of urine was carried out at 298 K on a Varian INOVA-600 spectrometer, operating at 600.13 MHz for  $^1\text{H}$  resonance frequency, equipped with a 5-mm inverse cryogenic probe. A standard one-dimensional pulse sequence [recycle delay (RD)-90 $^{\circ}$ - $t_1$ -90 $^{\circ}$ - $t_m$ -90 $^{\circ}$ ] was used to obtain urinary metabolite profiles with 90 $^{\circ}$  pulse length of  $\approx 10 \mu\text{s}$ . Water suppression was achieved with a weak irradiation equivalent to  $\approx 50 \text{ Hz}$  during the recycle delay (RD, 2 s) and mixing time ( $t_m$ , 100 ms). One hundred twenty-eight transients were collected into 32,768 data points for each spectrum with a spectral width of 12 kHz, and total repetition time was 3 s. An exponential window function with line-broadening factor of 1 Hz was applied to all free induction decays (FIDs) before Fourier transformation (FT). Metabolites concerned were identified according to literature data (9, 38) and selected metabolites were further

confirmed with a catalog of two-dimensional NMR experiments, including  $^1\text{H}$ - $^1\text{H}$  COSY,  $^1\text{H}$ - $^1\text{H}$  total correlation spectroscopy,  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear single-quantum correlation, and  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear multiple-bond correlation 2D NMR spectroscopy (see details in *SI Text: Materials and Methods*).  $^1\text{H}$  NMR spectra were manually corrected for phase and baseline and referenced to TSP ( $\delta$  0.0).

**Multivariate Data Analysis.** Initially, PCA was performed (on mean-centered data) to visualize the general structure of each dataset. Sex differences in  $^1\text{H}$  NMR and each DGGE dataset were investigated by OPLS-DA (18), where each data matrix was regressed against a dummy matrix of ones and zeros indicating sex class. The models were evaluated by assessment of the cross-validated scores from the model based on fivefold cross-validation (data were mean-centered and scaled to unit-variance before modeling). Sex-related variables were determined by interpretation of each OPLS-DA model.

Urinary metabolite profile and gut microbiota composition data were

- Lederberg J (2000) Infectious history. *Science* 288:287–293.
- Nicholson JK, Holmes E, Wilson ID (2005) Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol* 3:431–438.
- Nicholson JK (2006) Global systems biology, personalized medicine and molecular epidemiology. *Mol Syst Biol* 2:52.
- Eckburg PB, et al. (2005) Diversity of the human intestinal microbial flora. *Science* 308:1635–1638.
- Gill SR, et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312:1355–1359.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) Microbial ecology: human gut microbes associated with obesity. *Nature* 444:1022–1023.
- Martin FP, et al. (2007) A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. *Mol Syst Biol* 3:112.
- Li H, et al. (2007) Pharmacometabonomic phenotyping reveals different responses to xenobiotic intervention in rats. *J Proteome Res* 6:1364–1370.
- Wang Y, et al. (2004) Metabonomic investigations in mice infected with *Schistosoma mansoni*: an approach for biomarker identification. *Proc Natl Acad Sci USA* 101:12676–12681.
- Dumas ME, et al. (2006) Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc Natl Acad Sci USA* 103:12511–12516.
- Zoetendal EG (2001) The host genotype affects the bacterial community in the human gastrointestinal tract. *Microb Ecol Health Dis* 13:129–134.
- Turnbaugh PJ, et al. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444:1027–1031.
- Wang Y, et al. (2005) A metabonomic strategy for the detection of the metabolic effects of chamomile (*Matricaria recutita* L.) ingestion. *J Agric Food Chem* 53:191–196.
- Gavaghan CL, Holmes E, Lenz E, Wilson ID, Nicholson JK (2000) An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett* 484:169–174.
- Dumas ME, et al. (2006) Assessment of analytical reproducibility of  $^1\text{H}$  NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal Chem* 78:2199–2208.
- Schloss PD, Larget BR, Handelsman J (2004) Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl Environ Microbiol* 70:5485–5492.
- Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71:8228–8235.
- Trygg J, Wold S (2002) Orthogonal projections to latent structures (O-PLS). *J Chemometrics* 16:119–128.
- Mueller S, et al. (2006) Differences in fecal microbiota in different European study populations in relation to age, gender, and country: A cross-sectional study. *Appl Environ Microbiol* 72:1027–1033.
- Kelly D, et al. (2004) Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR- $\gamma$  and RelA. *Nat Immunol* 5:104–112.
- Xu J, et al. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* 299:2074–2076.
- Pryde SE, Duncan SH, Hold GL, Stewart CS, Flint HJ (2002) The microbiology of butyrate formation in the human colon. *FEMS Microbiol Lett* 217:133–139.
- Oh DK, et al. (2003) Production of conjugated linoleic acid by isolated Bifidobacterium strains. *World J Microbiol Biotechnol* 19:907–912.
- Selmer T, Andrei PI (2001) p-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycol radical enzyme catalysing the formation of p-cresol. *Eur J Biochem* 268:1363–1372.
- Acinas SG, Sarma-Rupavarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* 71:8966–8969.
- Felske A, Rheims H, Wolterink A, Stackebrandt E, Akkermans AD (1997) Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacteria in grassland soils. *Microbiology* 143(Pt 9):2983–2989.
- Muyzer G (1999) DGGE/TFGE a method for identifying genes from natural ecosystems. *Curr Opin Microbiol* 2:317–322.
- Margulies M, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
- DiCello F, et al. (1997) Biodiversity of a Burkholderia cepacia population isolated from the maize rhizosphere at different plant growth stages. *Appl Environ Microbiol* 63:4485–4493.
- Hayashi H, Sakamoto M, Benno Y (2002) Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiol Immunol* 46:535–548.
- Gonzalez JM, Zimmermann J, Saiz-Jimenez C (2005) Evaluating putative chimeric sequences from PCR-amplified products. *Bioinformatics* 21:333–337.
- Maidak BL, et al. (2001) The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* 29:173–174.
- Ludwig W, et al. (2004) ARB: A software environment for sequence data. *Nucleic Acids Res* 32:1363–1371.
- Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* 71:1501–1506.
- Muyzer G, de Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59:695–700.
- Pang X, et al. (2005) Molecular profiling of *Bacteroides* spp. in human feces by PCR-temperature gradient gel electrophoresis. *J Microbiol Methods* 61:413–417.
- Shen J, et al. (2006) Molecular profiling of the *Clostridium leptum* subgroup in human fecal microflora by PCR-denaturing gradient gel electrophoresis and clone library analysis. *Appl Environ Microbiol* 72:5232–5238.
- Tang H, Wang Y, Nicholson JK, Lindon JC (2004) Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. *Anal Biochem* 325:260–272.

**ACKNOWLEDGMENTS.** We thank Hui Dai for her assistance in solid-phase extraction work. This work was supported by 973 Program Grants 2003CB515506, 2004CB518600, 2006CB503909, 2007CB914701, and 2007CB513002; Shanghai Basic Research Program Grant 05DJ14009; National Natural Science Foundation of China Program Grant 30730005; Chinese Academy of Sciences 100T program Grant T12508-O6S138; and International Cooperation Program Grants 2007DFC30450 and 075407001. M.R. is supported by the METAGRAD program at Imperial College London.